NASA'S ADVANCED INFORMATION SYSTEMS TECHNOLOGY (AIST): COMBINING NEW OBSERVING STRATEGIES AND ANALYTICS FRAMEWORKS TO BUILD EARTH SYSTEM DIGITAL TWINS

Jacqueline Le Moigne

NASA Science Mission Directorate (SMD) Earth Science Technology Office (ESTO)

ABSTRACT

NASA's Advanced Information Systems Technology (AIST) Program is one of several Technology programs managed by the Earth Science Technology Office (ESTO) in the Earth Science Division (ESD). The AIST Program focuses on advanced information systems and novel computer science technologies that will be needed by NASA Earth Science in the next 5 to 10 years. The three main thrusts of the AIST Program deal with New Observing Strategies, Analytic Collaborative Frameworks and Earth System Digital Twins. This paper describes these three thrusts and how they work together to create future Earth Science information systems.

Index Terms— Earth Science Technology, Information Systems, New Observing Strategies, Analytic Collaborative Frameworks, Earth System Digital Twins

1. INTRODUCTION

As part of NASA Earth Science Technology Office, the Advanced Information Systems Technology (AIST) Program identifies, develops, and supports the adoption of software and information systems, as well as novel computer science technologies expected to be needed by the Earth Science Division in the 5-10-year timeframe. Projects under this Program start at a Technology Readiness Level (TRL, [1]) between 1 and 3 and usually advance one or two TRLs before completion. AIST information systems and software technologies contribute to the entire data lifecycle, as represented in Figure 1, from the acquisition of new measurements and the design of new observing systems to onboard intelligent data understanding and decision making and all the way to data analytics and extraction of the "science data intelligence" needed to create actionable information. Within that framework, AIST focuses on three thrusts:

(1) The first one, *New Observing Strategies (NOS)*, enables new observation measurements and new observing systems design and operations through intelligent, timely, dynamic, and coordinated distributed sensing. This contributes to the first part of the data lifecycle.

(2) The second thrust, *Analytic Collaborative Frameworks (ACF)*, enables agile science investigations that fully utilize the large amount of diverse observations using advanced analytic tools, visualizations, and computing environments. This addresses the second part of the data lifecycle.
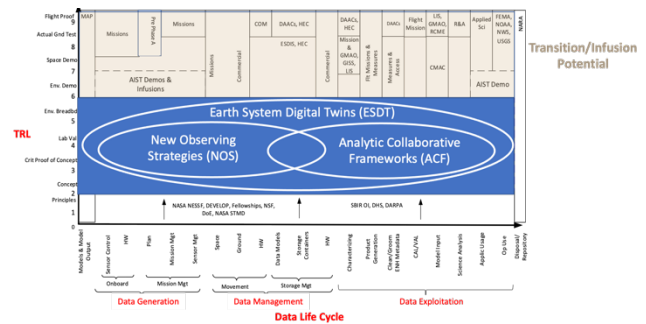

Figure 1 – *The AIST Program Spans the Entire Earth Science Data Lifecycle*

(3) The third thrust, *Earth System Digital Twins (ESDT)*, enables the development of integrated Earth Science frameworks that mirror the Earth with state-of-the-art models (Earth system models and others), timely and relevant observations, and analytic tools. These information systems can be used for supporting near- and long-term science and policy decisions. Here "science decisions" include planning for the acquisition of new measurements, development of new models or science analysis, integration of Earth observations in novel ways as well as various applications to inform choices, support decisions, and guide actions, e.g., related to climate change or for societal benefit. As described in Section 2.3 below, ESDT frameworks will build upon NOS and ACF technologies to integrate a continuous stream of observations, interconnected models, data analytics, data assimilation, simulations, advanced visualizations and the ability to conduct "what-if" scenarios, for example to assess the impact of human activity on natural phenomena. Therefore, technologies associated with ESDT contribute to the entire data lifecycle, as shown in Figure 1.

The following sections describe the three AIST thrusts, highlight the main technologies associated with these concepts, and then show a few AIST project examples.

2. AIST THRUSTS DESCRIPTION

This section gives more details about the three AIST thrusts.

2.1. New Observing Strategies (NOS)

NOS technologies concentrate on optimizing measurement acquisition by using diverse observing and modeling capabilities representing various resolutions, dynamically coordinated and collaborating to provide complete representations of Earth Science phenomena. The observing assets can be in space, in the air, or *in situ*, and the observed phenomena may exist on a variety of spatial or temporal scales (e.g., real-time tracking of hazards and disasters or long-term asset coordination for continuous ecosystem monitoring). NOS can be described as a federated Observing System, a generalized SensorWeb, or more generally as an "Internet-of-Space (IoS)" concept in which each node can be: a sensor; a group of sensors; a constellation of satellites (e.g., Earth System Observatory concept [2]); a model or integrated models; or even a database or any other source of relevant information. The nodes have varying degrees of coordination to achieve a common science objective. The two main objectives of the NOS thrust are to:

(1) Design and develop future observation concepts at the request of a new measurement, for example as identified in the latest Decadal Survey or as the result of a model or other science data analysis. This includes leveraging existing assets and potentially augmenting and complementing them with adequate observations, sensors and information in order to produce the desired measurement.
(2) Dynamically respond to science and applied science events of interest, not only focusing on rapid disaster-like events, but also considering mid- and long-term science events and various area coverages, from global to regional to local-impact events, (e.g., distressed vegetation, potential landslides due to runoff, etc.).

NOS use cases and technologies of interest are summarized in the 2020 NOS Workshop Report [3]. As part of the NOS focus, the AIST Program is also developing a NOS Testbed (NOS-T) to validate new NOS technologies, independently and as a system, as well as to evaluate new observing concepts and to increase the TRL level of these new NOS technologies. Proof-of-concept demonstrations have been conducted using the prototype NOS-T and are reported in [4,5]. Figure 2 represents the AIST NOS concept.
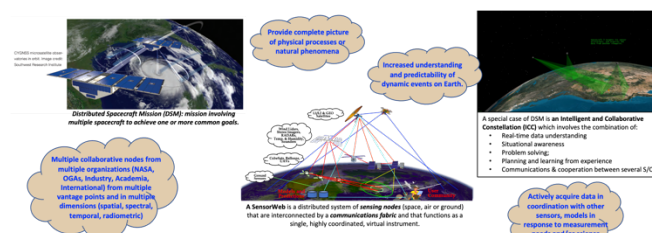

Figure 2 – *The AIST New Observing Strategies (NOS) Concept*

2.2. Analytic Collaborative Frameworks (ACF)
ACF technologies address the challenges associated to observing systems such as NOS systems that will acquire an increased variety and volume of data over various geographical scales, latencies, and frequencies. The ACF thrust is designed to facilitate access, integration, and understanding of large amounts of disparate datasets. Its purpose is to harmonize analytics tools, data, visualization and computing environments to meet the needs of Earth science investigations and applications. The ACF thrust integrates new or previously unlinked datasets, tools, models, and a variety of computing resources together into a common platform to address previously intractable scientific and science-informed application questions. Additionally, this activity seeks to generalize custom or unique tools that are currently used by a limited community of experts or practitioners, to make them accessible and useful to a broader community. ACF focus on reducing the amount of time a science user spends on data preparation and enable the tailoring of configurations of datasets and reusable tools to avoid repetitive work (e.g., by developing reusable components).

Both NOS and ACF aim at optimizing Earth Science mission return – NOS from an observation point of view and ACF from an analysis point of view, with assets and data from NASA and non-NASA sources. Additionally, with NOS dynamic capabilities and the ability of ACF systems to determine which additional observations would be needed, a feedback loop from ACF to NOS can be created with observation requests being sent from models to agile observing systems. As a result, these integrated, autonomous, distributed, and adaptable systems can be used as building blocks towards future ESDT systems described in section 2.3.

2.3. Earth System Digital Twins (ESDT)
The Digital Twin concept was defined in 2002 [6] and was first developed for industry for Product Lifecycle Management purposes. Its goal was to create a digital information system that would be a "twin" of the physical system through its entire lifecycle, utilizing both the initial construct of that system as well as all the continuous information available from sensors embedded within the physical system. Over time the concept has been applied to other domains such as infrastructure development and cars, and more recently to the human body and to the Earth System [7]. Depending on the domain, various definitions have been proposed [8]; the one proposed in [7] for ESA's Destination Earth project is: "A digital twin of Earth is an information system that exposes users to a digital replication of the state and temporal evolution of the Earth system constrained by available observations and the laws of physics."

The AIST Program defines an *Earth System Digital Twin (ESDT)* as an interactive and integrated multidomain, multiscale, digital replica of the state and temporal evolution of Earth systems, that dynamically integrates: relevant Earth system models and simulations; other relevant models (e.g., related to the world's infrastructure); continuous and timely (including near real time and direct readout) observations (e.g., space, air, ground, over/underwater, Internet of Things (IoT), socioeconomic); long-time records; as well as analytics

and artificial intelligence tools. Effective ESDTs enable users to run hypothetical scenarios to improve the understanding, prediction of and mitigation/response to Earth system processes, natural phenomena and human activities as well as their many interactions.

With this definition, we can see how an ESDT is a type of integrated information system that, for example, enables continuous assessment of impact from naturally occurring and/or human activities on physical and natural environments. AIST ESDT strategic goals are to:

1. Develop information system frameworks to provide continuous and accurate representations of systems as they change over time;
2. Mirror various Earth Science systems and utilize the combination of Data Analytics, Artificial Intelligence, Digital Thread and state-of-the-art models to help predict the Earth's response to various phenomena;
3. Provide the tools to conduct "what if" investigations that can result in actionable predictions.

From an AIST point of view, ESDT capabilities integrate Earth observations analysis and understanding capabilities provided by ACF-type systems and on-demand and timely IoT and IoS data using NOS capabilities, while taking advantage of advanced Machine Learning, Big Data Analytics, and powerful computational and visualization capabilities.

The AIST ESDT thrust is also investigating how several local and/or thematic Digital Twins built at multiple spatial and temporal scales can be federated in a hierarchical fashion to construct a global Digital Twin of the Earth. For such a vision, the first step will be to define interoperability and interface standards. Figure 3 illustrates the complexity of ESDT-like systems.
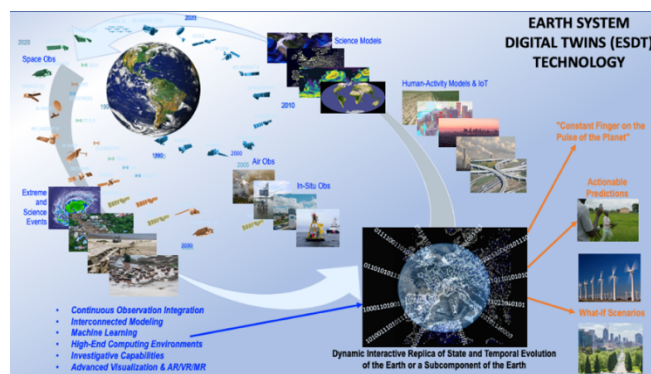

Figure 3 – *Future AIST Earth System Digital Twins (ESDT) Information Systems will Integrate Many Various Components*

3. AIST TECHNOLOGIES

In order to successfully develop these future complex information systems, many existing capabilities will need to be integrated and new ones will need to be developed. Following are some of the most important capabilities that will be required for each of the thrusts defined earlier:

- *NOS systems capabilities*: Semantic representations, ontologies, and AI- and knowledge-based agents used to represent disparate nodes of NOS/IoS systems; technologies for onboard high-level autonomous decision making, including deliberative/deduction AI (e.g., heuristic search, probabilistic reasoning) and Hybrid AI (reflexive plus deliberative); and edge intelligence for faster data-driven decision making.
- *ACF framework capabilities*: Multi-disciplinary ACF concepts (e.g., enabling analysis of cascading impacts); analytics and workflow management tools capable of characterizing natural phenomena or physical processes from extremely large amounts of diverse data and information, including non-traditional and unstructured data (e.g., IoT, socioeconomic, social media); and data-driven modeling tools enabling the forecast of future behavior of Earth Science phenomena.
- *ESDT capabilities*: Technologies for agile interaction and interoperability between measurement acquisition (NOS or NOS-like) and science investigations (ACF or ACF-like); frameworks that enable data ingest from multiple, integrated models; technologies for developing "federated ESDTs"; novel AI techniques enabling faster data processing, fusion and analysis as well as for surrogate modeling at multiple scales that will facilitate "what-if" investigations inherent to ESDT systems; and innovative visualization methods such as Mixed Reality (MR) techniques.

4. EXAMPLES OF AIST PROJECTS

All AIST projects address various aspects of the technology needs described in Section 3. Tables 1 and 2 below show a few current projects representing NOS and ACF thrusts.

Table 1 - A Few AIST NOS Current Projects

PI's Name/Org	Title	Synopsis
Moghaddam/U. of Southern California	SPCTOR: Sensing Policy Controller and OptimizeR	Multi-sensor coordinated operations and integration for soil moisture, using ground-based and UAVs "Sensing Agents".
Carr/Carr Astro	StereoBit: Advanced Onboard Science Data Processing to Enable Future Earth Science	SmallSat/CubeSat high-level onboard science data processing demonstr. multi-angle imagers, using SpaceCube processor and CMIS Instrument, and Structure from Motion (SfM).
Nag/NASA ARC	D-SHIELD: Distributed Spacecraft with Heuristic Intelligence to Enable Logistical Decisions	Suite of scalable software tools - Scheduler, Science Simulator, Analyzer to schedule payload ops of large constellation based on constraints, resources, and subsystems.
Grogan/Stevens Inst. of Technology	Integrating TAT-C, STARS, and VCE for New Observing Strategy Mission Design	Design of Pre-Phase A distributed space mission concept, by integrating architecture enumeration, high-level evaluation, adaptive sensor interaction & onboard computing/networking
Posselt/JPL	Parallel OSSE Toolkit	Fast-turnaround, scalable OSSE Toolkit to support both rapid and thorough exploration of the trade space of possible instrument configurations
Forman/Univ. of Maryland	Next Generation of Land Surface Remote Sensing	Terrestrial hydrology OSSE/mission planning tool for terrestrial snow, soil moisture, and vegetation using passive/active microwave RS, LiDAR, passive optical RS, hydrologic modeling, and data assimilation
Gutmann/UCAR	Future Snow Missions: Integrating SnowModel in LIS	Improve NASA modeling capabilities for snow OSSE, for a future cost-effective snow mission by coupling the SnowModel modeling system into NASA LIS.

With respect to ESDT, the first prototype is being developed as a collaboration between the AIST Program and CNES under the umbrella of the Space Climate Observatory (SCO). The Integrated Digital Earth Analysis System

(IDEAS)/FloodDAM DT is an extension of the previous FloodDAM flood analysis and prediction system [9]. The IDEAS project will bridge the physical environment and its virtual representation by continuously assimilating new observations and will leverage several advanced nymerical models and analysis. IDEAS/FloodDAM DT will enable scenario-based flood prediction for assessing infrastructure and population impacts. More information about IDEAS can be found in [10].

5. CONCLUSION

An overview of NASA's Earth Science Advanced and Information Systems Technology (AIST) Program was presented, including its three major thrusts, New Observing Strategies (NOS), Analytics Collaborative Frameworks (ACF) and Earth System Digital Twins (ESDT). After integration, these AIST capabilities and projects will form the basis towards the development of future Earth Science information systems. Additional AIST project examples and details can be found on the AIST website [11].

6. ACKNOWLEDGEMENTS

The author would like to thank the AIST team for many stimulating discussions as well as ESTO for its support.

7. REFERENCES

[1] "Definitions of Technology Readiness Levels," 01/14/22, https://esto.nasa.gov/files/trl_definitions.pdf.

[2] "NASA Earth System Observatory (ESO)," Updated: 01/15/22, https://science.nasa.gov/earth-science/earth-system-observatory.

[3] J. Le Moigne and M. Cole, "AIST NOS Workshop Summary Report," https://esto.nasa.gov/wp-content/uploads/2021/02/AIST-NOS-Workshop-Report.pdf, February 2020.

[4] B. Smith et al, "Demonstrating a New Flood Observing Strategy on the NOS Testbed," *IGARSS'2022*, July 22-27, 2022, Malaysia.

[5] M.S. Seablom et al, "2022 New Observing Strategies Live Demonstration," *IGARSS'2022*, July 22-27, 2022, Malaysia.

[6] M. Grieves and J. Vickers, "Digital Twin: Mitigating Predictable, Undesirable Emergent Behavior in Complex Systems," in *Transdisciplinary Perspectives on Complex Systems*, Eds. F-J. Kahlen et al, Springer, 2017.

[7] P. Bauer et al, "A Digital Twin of Earth for the Green Transition," *Nature Climate Change*, Vol. 11, Feb.2021, pp. 80-83.

[8] "Glossary of Digital Twins", the Digital Twin Consortium, 2022, https://www.digitaltwinconsortium.org/glossary/index.htm.

[9] P. Kettig et al, "The SCO-FloodDAM Project: New Observing Strategies for Flood Detection, Alert and Rapid Mapping," IGARSS'2021, July 11-16, 2021, Brussels, Belgium.

[10] T. Huang et al, "An Earth System Digital Twin for Flood Prediction and Analysis," *IGARSS'2022*, July 22-27, 2022, Malaysia.

[11] Advanced Information Systems Technology (AIST) Program, https://esto.nasa.gov/aist/.

Table 2 - A Few AIST ACF Current Projects

PI's Name	Title	Synopsis
Schollaert Uz/NASA GSFC	Supporting shellfish aquaculture in the Chesapeake bay using AI for water quality	Access to reliable information on a variety of environmental factors, not currently available at optimal scales in space and times, by using various data and AI for Pattern Recognition.
Moisan/NASA GSFC	NASA Evolutionary Programming Analytic Center (NEPAC)	Discover novel algorithms for ocean chlorophyll using Genetic Programming on satellite/in-situ obs connecting data and applications with HEC resources.
Jetz/Yale U.	Biodiversity - Environment Analytic Center	Near real-time monitoring of Earth's biological pulse/multi-scale analysis, visualization and change detection; online dashboard, various res, data uncertainty & biodiv. databases.
Townsend/U Wisdonsin, Madison	GeoSPEC: On-Demand Geo-spatial Spectroscopy Proc. Environment on the Cloud	Framework/processing workflow for on-demand cloud-based Hyperspectral/ Spectroscopy Science Data Processing.
Swenson/ Duke Univ.	Canopy condition to continental scale biodiversity forecasts	Characterize canopy condition from various spatio-temporal RS products to predict supply of mast resources to herbivores and visualize canopy/drought-stress maps
Ives/U. of Wi, Madison	Valid time series analyses for satellite data	New statistical tools to analyze large, time series of various remotely sensed datasets with improved statistical rigor and confidence. Applied to change patterns.
Martin/Washington U.	Development of GCHP to enable broad community access to high-resolution atmospheric composition modeling	Integrate atmospheric chemistry models online into ES models and offline using meteorological data, high-perf. version of GEOS-Chem and the Earth System Modeling Framework (ESMF).
Duren/NASA JPL	Multi-scale Methane Analytic Framework	ACF for methane data analysis spanning multiple observing systems and spatial scales with workflow optimization, analytic tools tools for data search and discovery, and a collaborative, web-based portal.
Henze/U. of CO, Boulder	Surrogate modeling for atmospheric chemistry and data assimilation	Improved computational tools for AQ prediction, mitigation, research; computationally efficient chemical data assimilation; compressive sampling and ML for large-scale dynamical syst; multi-source data.
Holm/City of Los Angeles	Predicting What We Breathe: Using Machine Learning to Understand Urban Air Quality (AQ)	Link ground-based in situ and space-based remote sensing obs to classify patterns in urban AQ, predict air pollution events, identify similarities between megacities, using science models and ML-based algs.
Beck/U. of AL, Huntsville	Cloud-based Analytic Framework for Precipitation Research	Cloud-based ACF for Precipitation Research using a Deep Learning framework; analysis-optimized cloud data store and on-demand cloud-based serverless tools.
Coen/NCAR	Creation of a Wildfire Fire Analysis: Products to Enable Earth Science	Create, test, assess wildland fire reanalysis products using fire detection data, as well as coupled weather-wildland fire model and data assimilation.
Donnellan/ NASA JPL	Quantifying Uncertainty and Kinematics of Earthquake Systems ACF (QUAKES-A)	Info Services & uniform crustal deformation reference model for active plate margin of CA fusing widely varying spatial and temporal res data + quantifying uncertainty.
Hua/NASA JPL	Smart On-Demand of SAR ARDs in Multi-Cloud & HPC	Full resolution time series analysis, high-accuracy flood and damage assessments with RS SAR Analysis Ready Data (ARD) (Jupyter Notebooks and on-demand analysis multi-cloud environments).
Huffer/ Lingua Logica	AMP: An Automated Metadata Pipeline	Automate/improve use and reuse of Earth Science data with fully-automated metadata semantic mining pipeline integrating ML and SWEET ontologies.
Zhang/ Carnegie Mellon U.	Mining Chained Modules in Analytics Center Framework	Workflow tool recommending SW modules. Leverages Jupyter Notebooks to mine SW usage history, extract reusable module chains, & develop intelligent service.